THE UNIVERSITY of York

Discussion Papers in Economics

No. 2006/23

The 'weight of evidence' concerning tobacco harm: beliefs in mid-twentieth century America

by

Martin Forster, Martin Walsh and Sue Bowden

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

The 'weight of evidence' concerning tobacco harm: beliefs in mid-twentieth century America

Martin Forster* Martin Walsh† Sue Bowden*

December 7, 2006

Department of Economics and Related Studies

Discussion paper no. 06/23

Abstract

We present a version of Chern et al.'s (1995) Bayesian model of 'health risk belief' to track the evolution of the 'weight' of epidemiological evidence concerning tobacco harm that was in the possession of the U.S. Tobacco Industry Research Committee (T.I.R.C.) and the U.S. Public Health Service and related groups during the 1950s and early 1960s. We compare our results with public statements assessing the evidence that were made by the organisations during the same period. The results from the models for the U.S. Public Health Service and related groups are not in disagreement with the public statements of these organisations; the results from the lung cancer model for the T.I.R.C. are in disagreement with the assessments of the evidence made by the T.I.R.C.'s Scientific Director in his annual reports. We discuss possible reasons for this, relating our findings to present-day academic and legal debates about the 'controversy' surrounding tobacco harm during the mid-twentieth century.

dp 07/12/06

*Department of Economics and Related Studies and Centre for Historical Economics and Related Research at York, University of York, Heslington, York YO10 5DD, U.K.. Corresponding author: Martin Forster (e-mail: mf8@york.ac.uk, tel.: +44 (0)1904 433797; fax +44 (0)1904 433759).

†Department of Economics and Related Studies, University of York.

1 Introduction

Tobacco-related lawsuits worldwide have expended vast amounts of resources arguing about the tobacco industry's role in the 'controversy' surrounding tobacco use and ill health during the mid-twentieth century. Central to these debates has been the issue of when the 'weight' of scientific evidence indicting tobacco became sufficient to declare its use harmful to health.

In this paper we present a version of Chern et al.'s (1995) Bayesian model of 'health risk belief' to track the evolution of the 'weight' of epidemiological evidence concerning the relationship between tobacco use and six cancer sites that was in the possession of a number of American organisations during the 1950s and early 1960s: the Tobacco Industry Research Committee (T.I.R.C., renamed the Council for Tobacco Research-U.S.A., Inc., (C.T.R.) in 1964), an industry body, and the U.S. Public Health Service and related groups. Our sources come from a recent review of the epidemiological literature collected by these organisations (Forster et al. (2006)). To assess the 'fit' of the models, we compare the results with public statements made by the organisations during the same period. The results are also used to lend insight into the manner in which the epidemiological evidence suggesting that tobacco use is harmful accumulated during the early- to mid-twentieth century.

We adopt a 'qualitative', Bayesian, perspective to the assimilation of evidence, rather than the modern-day approach of cumulative meta-analysis. We argue that our approach is better-suited to the way in which the T.I.R.C. was evaluating the articles within its possession and can also be used to represent elements of the way in which the U.S. Public Health Service and related groups assessed the evidence. The paper complements recent work by Parascandola et al. (2006) who present, for the U.S. Public Health Service, a review of the manner in which two U.S. Surgeon General's reports into smoking and health applied causal criteria to the epidemiological evidence that they reviewed. We find that the results of our models for the U.S. Public Health Service and related groups do not contradict their public assessments of the evidence, but that the results of the models for the T.I.R.C. do not fit well the opinions expressed in the annual reports of its Scientific Director, C.C. Little. We discuss possible reasons for this to be the case, including the role of prior beliefs, the weight given by each organisation to epidemiological evidence in an overall assessment of causality and the conflicting objectives of the U.S. Public Health Service and related groups and the T.I.R.C..

The themes of this paper - an evolving evidence base suggesting harm, conflicting expert opinion and the involvement of state and industry - are reflected in other controversies, both contemporary and historical, surrounding harm to human health and/or the environment (see, for example, the European Environment Agency's (2001) 'Late Lessons from Early Warnings'). In such instances, policy makers, industry and wider society must try to as-

sess the state of evidence as it evolves over time and decide on what action, if any, they should take. For a policy-maker, this might involve regulation or prohibition; for the industry, changes in product specification to reduce risk; for the consumer, reducing consumption of, or ceasing to consume, the product.

This paper is one of a series of three to result from our research, the others being the review of the literature collected by the U.S. Public Health Service and related groups and the T.I.R.C., which provides the data for the inputs to the models in this paper (Forster et al., 2006), and a review of the inner-workings of the T.I.R.C. between 1954 and 1964, especially the role of the academics who served on its Scientific Advisory Board (Bowden et al., 2006).

The rest of this paper is organised as follows. In section 2 we discuss the historical, legal and academic background to the work before outlining the model (section 3), presenting results (section 4) and discussing them (section 5). Section 6 concludes.

2 Background

The invention of the 'Bonsack' machine in the 1880s allowed cigarettes to be produced at a fraction of the unit cost prevailing at the time, heralding an era in which the smoking habit became accessible to the whole of society rather than just those in elite circles (Hilton, 2000).[1] By 1955 in the U.S.A., around 65% of men and 35% of women were thought to be regular cigarette smokers (Haenszel et al., 1956).

It was concern with the increasing incidence of cancer, especially of the lung and bronchus, during the first half of the twentieth century, which led researchers to investigate whether tobacco use played a causal role. Retrospective epidemiological studies can be traced back to the early 1920s and large prospective studies to the 1950s.

The 1950s was the decade during which the evidence suggesting that cigarette smoking caused lung cancer started to cause serious concern within society. That decade was also the one in which the tobacco companies, warehouse and growers' associations formed the T.I.R.C. and the U.S. Public Health Service and related groups started to issue their first public statements assessing the evidence concerning smoking and lung cancer. During this decade and the next, much scientific argument took place about whether or not tobacco played a causal role in cancer, cardiovascular disease and other respiratory diseases, argument which has been revisited in the tobacco litigation cases and academic literature of today.[2]

[1] Parts of this section are based on the historical notes in Forster et al. (2006).

[2] More details of the historical context may be found in many sources, including Talley et al. (2004), Parascandola (2004) and Doll (1998).

2.1 Contemporary legal and academic perspectives

Recent legal judgments in tobacco-related lawsuits in the U.S.A., the United Kingdom and the Republic of Ireland have expressed markedly differing opinions on dating the end of the 'controversy' surrounding whether or not tobacco is harmful to health. Their verdicts and orders have been equally varied. The 2006 Racketeer Influenced and Corrupt Organizations (R.I.C.O.) judgment (United States of America v. Philip Morris U.S.A., Inc., et al., 2006) concluded that the 'overwhelming medical and scientific consensus' that smoking cigarettes causes disease had been established by the early 1950s [page 3] and that the tobacco companies had been involved in 'a massive 50-year scheme to defraud the public' [page ES-1]. It judged the United States to be entitled to 'at least' $280bn of 'disgorgement of proceeds' [page ES-26]. Also in 2006, the Court of Appeal of the State of California cited the late 1950s as being the time by which there existed general agreement that smoking caused lung cancer within the scientific and medical profession 'after several epidemiological studies reached that conclusion' [page 3]. It affirmed an award of $28m in punitive damages against Philip Morris U.S.A., Inc., for its 'extremely reprehensible' conduct in perpetuating a 'false controversy' regarding tobacco's harmful effects during the 1950s and beyond (pages 2, 15, 65, Jodie Bullock v. Philip Morris, U.S.A., Inc., 2006).[3]

However, judgments on the other side of the Atlantic have reached markedly different conclusions. In 2005, in the Scottish Court of Session, Lord Nimmo Smith judged that, by 1964, the general public in the United Kingdom were 'well aware' of the view that smoking could cause lung cancer [par. 9.4]. However, the defendant, Imperial Tobacco Limited, averred that '[c]igarette smoking has not been scientifically established as a cause of lung cancer' and that 'the cause or causes of lung cancer are unknown and the mechanism or mechanisms whereby lung cancer develops are unknown' [par. 2.7]. Lord Nimmo Smith judged it was not within 'judicial knowledge' that smoking caused lung cancer [par. 9.7] and that the plaintiff, a victim of lung cancer, had failed to prove that his lung cancer was probably caused by his smoking [par. 6.185]. Nimmo Smith dismissed the plaintiff's claim for damages (McTear v. Imperial Tobacco Ltd., 2005). In 2004 in the High Court of Ireland, Justice F. Geoghegan dismissed claims of alleged negligence by three plaintiffs against three tobacco companies without assessing any evidence. In part of his judgment he commented:

> 'the claims would require the court to decide issues of fact pertaining to the state of scientific knowledge [regarding tobacco harm] which they either were aware of or ought to have been

[3] Opinion is presently noted as being 'superseded'. A petition for review briefing is deferred pending a decision on the level of punitive damages to be awarded in Philip Morris U.S.A. v. Williams or further order of the court. See `www.courtinfo.ca.gov/opinions/archive/B164398.PDF`

> aware of and the precise decision taken by the defendants not only in relation to the manufacturing but including detailed decisions effecting such matters as the level of nicotine over much of the twentieth century . . . For a court to be asked in the years 2006 or 2008 or later to determine issues of fact of the nature which would be required by these claims. . . "puts justice to the hazard"' (Section entitled 'Dismiss in the interests of justice', Manning v. Benson and Hedges Ltd., 2004).

Such debates are not limited to courts of law; the subject has also exercised many in the academic world. Talley et al. (2004) argue that 'there was a legitimate and reasonable scientific controversy over cigarette smoking and lung cancer in the 1950s and early 1960s' [page 329]. They describe Stolley's (1991), Kluger's (1997) and Glantz et al.'s (1996) accounts of the scientific controversy in the mid-20th century as being 'ahistorical' in not having adopted an appropriate historical perspective. Parascandola (2004) writes about the roles of statisticians R. A. Fisher and J. Berkson, both of whom, he notes, have been accused of 'taking the 'wrong' side of the debate' in the controversy. He argues that Fisher's and Berkson's support for the controlled experiment, the 'crucial, objective test of a causal hypothesis', was in conflict with the views of those who argued for an 'inferential judgment based on a diverse body of evidence.' Parascandola cites the Surgeon General's 1964 report as being the 'authoritative statement of the proper method for assessing a body of etiological evidence.' Such legal and academic debate is not helped by the potential conflicts of interest thrown up by tobacco industry funding of scientists and academics as consultants and expert witnesses both today and during the 1950s and 1960s.[4]

[4]In a declaration fronting Talley et al. (2004) it is stated that two of the authors (Talley and Kushner) have consulted with a tobacco industry law firm specialising in 'recruiting and developing expert witnesses for the defense in tobacco industry lawsuits.' Kushner is described as having ceased ties with the firm and it is explained that neither author plans to work on tobacco-related litigation in the future. To support their argument about the legitimacy of the 'controversy' prior to 1964, Talley et al. cite the views of Paul M. Kotin [pages 349, 351, 354], who served as a member of the Scientific Advisory Board of the T.I.R.C., between 1954 and 1965 (when he resigned). This membership is not stated in Talley et al.'s paper. Talley et al. and Parascandola also cite the views of Berkson who, according to the recent R.I.C.O. judgment (pages 64, 286, United States of America v. Philip Morris U.S.A., Inc., et al., 2006), at the time of publication of the Surgeon General's 1964 report 'Smoking and Health', was receiving consultancy fees of $8,000 per annum from 'Special Account No. 3' operated by Philip Morris, R.J. Reynolds, Lorillard, Liggett, Brown and Williamson and American Tobacco. According to the R.I.C.O. judgment, Special Account No. 3 was created by the tobacco companies to be a tobacco lawyers' 'work product' and thus not subject to subpoena (page 203, United States of America v. Philip Morris U.S.A., Inc., et al., 2006). It is not clear to us when this information regarding Berkson's consultancy became public. Neither Talley et al. nor Parascandola make reference to it. Berkson's arguments formed part of the defence case for Imperial Tobacco Ltd. in McTear v. Imperial Tobacco Ltd. (2005).

2.2 Roles of the T.I.R.C. and the U.S. Public Health Service and related groups

The Tobacco Industry Research Committee

The T.I.R.C. was founded at the end of 1953 by the majority of the U.S. tobacco companies, growers and warehouse associations. It aimed to 'aid and assist research into tobacco use and health, and particularly into the alleged relationship between tobacco use and lung cancer, and to make available to the public factual information on this subject' (page 3, C.T.R., not dated). In an attempt to split the research and public relations aspects of its work, a 'Scientific Advisory Board' was created and charged with promoting and funding scientific research. Also formed were a Law Committee and an Industry Technical Committee, which comprised the research directors of the member companies (pages 3, 5, C.T.R., not dated). The T.I.R.C. also operated a library to collect together papers investigating the relationship between tobacco and health, which served the Committee itself, its scientific staff, the S.A.B. and the tobacco industry. The library did not serve the general public (pages 7-8, C.T.R., not dated).

Documents from the T.I.R.C. archives have been made publicly available courtesy of what is known as the 'Master Settlement Agreement' of 1998, an agreement signed by the Attorneys General of 46 U.S. states and the four largest U.S. tobacco companies (National Association of Attorneys General, 1998). The agreement introduced major restrictions on the advertising, marketing and promotion of cigarettes, dissolved the C.T.R. and required the tobacco companies to open, at their own expense, websites holding all 'non-privileged' documents requested in relevant state and other smoking-related lawsuits (non-privileged documents being those documents for which the tobacco companies did not make a claim of attorney-client privilege, trade-secret protection and so on). These documents are collected together in a number of websites, including the Legacy Tobacco Documents Library at the University of California, San Francisco[5] and 'tobaccodocuments.org'.[6]

As described in detail in Forster et al. (2006), two confidential publications produced by the T.I.R.C.'s library which summarise literature on tobacco and health are used to provide the data for the T.I.R.C. models presented in this paper. One is entitled 'A working reference catalog' (hereafter 'the Catalog'), a T.I.R.C. listing of scientific literature on tobacco and health published prior to 1st August 1955. The other is entitled 'Current Digest of Scientific Papers Relating to Tobacco Use' (hereafter 'Current Digest'), a monthly publication first published in July 1956, which summarised the tobacco-health literature as it was published.[7]

[5] `http://legacy.library.ucsf.edu/`

[6] `http://tobaccodocuments.org/`

[7] A copy of the catalog may be obtained from `http://tobaccodocuments.org/ness/5305.html` or

The Catalog was compiled by the T.I.R.C.'s first librarian, Kenneth Austin. Austin obtained relevant scientific articles published prior to 1st August 1955 and classified them by author under a range of headings relating to the smoking/health question. Austin assigned a code letter A to E to each article that he summarised. 'A' meant that the paper took a 'favorable', 'B' a 'neutral or unrelated' and 'C' an 'unfavorable' position on the effects of tobacco (T.I.R.C., not dated (b)). 'D' denoted that '[n]either paper nor excerpt is on file, but copies ... can be obtained from medical libraries or publishers' and 'E' denoted '[c]opies of full paper on file but no excerpt or abstract available.' In a trial deposition from 1986, Austin's successor, William D. Jenkins, explains that he interpreted the letters A to C as follows: 'A' means 'not harmful ... tobacco is not harmful'; 'B' means 'neutral' and 'C' means 'tobacco is harmful' (page 42, deposition of W. D. Jenkins, Almquist et al. v. American Brands, Inc., et al., 1986).

To identify literature reviewed by the T.I.R.C. after 1st August 1955, we used the Current Digest. Austin scanned the scientific publications to which the T.I.R.C. subscribed, together with abstracts of the world literature and, together with his staff, checked the scientific publications at the New York Academy of Medicine. 'Pertinent' articles were then selected and used to compile the Current Digest. By 1968 it was claimed that the library received 129 scientific publications, ten publications abstracting the literature, and information on monitoring of 2,500 U.S. and foreign journals carried out by the Philadelphia College of Physicians (Austin, 1968).

Forster et al. (2006), in reviewing the quality of the T.I.R.C.'s literature retrieval and review process, found it to be unbiased and reasonably comprehensive. From 1956 onwards, it was providing summaries of published epidemiological studies to T.I.R.C. members within an average of 2.3 months of date of publication. It is the simple categorical classifications of the studies within the T.I.R.C.'s possession (tobacco harmful, not harmful, neutral) that are reported in Forster et al. (2006) that are used as inputs to the T.I.R.C. model presented in this paper.

Our sources for the T.I.R.C.'s public statements assessing the evidence are the annual reports of its Scientific Director, C.C. Little. Each report contains a general assessment of the evidence concerning tobacco harm and then presents detailed information on the research program and projects funded by the T.I.R.C.. In these reports, we focus on Little's statements concerning the role of epidemiological evidence in making judgments about causality and whether or not he believed tobacco use to be causally related to the cancer sites considered in this paper.

`http://legacy.library.ucsf.edu/tid/ejx71e00`. Copies of the Current Digest may be downloaded from the Legacy Tobacco Documents Library site.

The reports of the U.S. Public Health Service and related groups

The United States Public Health Service and related groups published a number of reviews of the literature and statements on smoking and lung cancer during the 1950s and a major report on tobacco use and health by an advisory committee to the U.S. Surgeon General in 1964.

In 1957, a scientific study group examined 18 independent epidemiological studies into the smoking/lung cancer question, together with experimental and pathological evidence, and concluded that '[t]he sum total of scientific evidence establishes beyond reasonable doubt that cigarette smoking is a causative factor in the rapidly increasing incidence of human epidermoid cancer of the lung' (page 1129, Study Group on Smoking and Health, 1957). Around the same time, Surgeon General Burney stated that the U.S. Public Health Service 'feels that the weight of evidence is increasingly pointing in one direction: that excessive smoking is one of the causative factors in lung cancer' (page 44, Burney, 1958).

In 1959, reviewing additional evidence, including newly-published epidemiological studies, Burney stated that 'the weight of evidence at present implicates smoking as the principal factor in the increased incidence of lung cancer . . . cigarette smoking particularly is associated with an increased chance of developing lung cancer' (pages 1835-6, Burney, 1959).

The Surgeon General's 'Advisory Committee on Smoking and Health' was an expert committee formed in 1962 and charged with making 'an objective assessment of the nature and magnitude of the health hazard [associated with tobacco use].' Its job was to review critically all available data but not to conduct new research. The Committee set out to investigate whether or not the use of tobacco was bad, good or devoid of effects on health and considered the question from the perspective of overall mortality as well as disease categories.

The Advisory Committee combined epidemiological evidence with that from experimental, clinical and pathological studies in an attempt to make judgments on causality for six cancer sites, cardiovascular disease and other diseases thought to be related to smoking. The Committee was clear about the value of epidemiological evidence, which was 'used extensively in the assessment of causal factors in the relationship of smoking to health among human beings upon whom direct experimentation could not be imposed.' It is the prospective and retrospective epidemiological studies which form the focus of this paper, that were described by the Committee as having furnished information 'of the greatest value.'

In contrast to the T.I.R.C., the Advisory Committee laid out explicit criteria for assessing the evidence within its possession. Further, it was clear about its view of the role of statistical associations in assessing causality: '[s]tatistical methods cannot establish proof of a causal relationship in an association. The causal significance of an association is a matter of judgment

U.S. Public Health Service and related groups

- **1957**: Study Group on Smoking and Health (1957); Surgeon General Burney's statement of 1957 (Burney, 1958)
- **1959**: Surgeon General Burney's statement of 1959 (Burney, 1959)
- **1964**: U.S. Surgeon General's Advisory Committee on Smoking and Health (U.S. Department of Health, Education and Welfare, Public Health Service, 1964)

Tobacco Industry Research Committee

- **Pre-August 1955**: the Catalog
- **Post-July 1956**: the Current Digest

FIGURE 1: Sources of data for Bayesian models solved in this paper.

which goes beyond any statement of statistical probability' [page 20]. The first stage in assessing the evidence is described by the Advisory Committee as establishing whether an association is observed between host and agent [page 20]. The second stage is assessing whether identified associations are causal [page 21]. Causality is taken to mean 'a significant, effectual, relationship between an agent and an associated disorder or disease in the host' [page 21]. The criteria for assessing causality included the consistency, strength and specificity of associations, as well as their temporal relationship and coherence (whether the results agree with the natural history and biology of the disease). These are described in more detail in the report itself, as well as in Forster et al. (2006) and Parascandola et al. (2006).

The Advisory Committee's report - 'Smoking and Health' - was published on January 11th, 1964. Finding cigarette smoking associated with a seventy per cent increase in the age-specific death rates of males and an increased death rate of females, it concluded that cigarette smoking 'contributes substantially to mortality from certain specific diseases and to the overall death rate' and that it 'is a health hazard of sufficient importance in the United States to warrant appropriate remedial action' [page 31]. Regarding the relationship between smoking and lung cancer, it concluded that '[c]igarette smoking is causally related to lung cancer in men; the magnitude of the effect of cigarette smoking far outweighs all other factors. The data for women, though less extensive, point in the same direction' [page 31-3].

A summary of the organisations and sources whose data form the inputs to our Bayesian model is shown in figure 1.

3 Application of Chern et al.'s model

Chern et al. (1995) model how a 'representative consumer's' beliefs regarding the health risk associated with saturated fat consumption evolved as the evidence base concerning the potential of saturated fats to harm human health changed between 1966 and 1987. The inputs to their model are the data from Brown and Schrader's (1990) 'cholesterol index', constructed from a review of medical journal publications which deal with the 'human and clinical implications' of cholesterol and which, in the opinion of Brown and Schrader, are most likely to have been read by U.S. physicians.[8] 935 articles are included in the series and classified by the authors as either supporting or attacking the hypothesised link between cholesterol and heart disease.

In the models presented in this paper, we consider the beliefs of a decision-maker who reviews a sequence of articles published in primary sources and which report the results of epidemiological studies (both prospective and retrospective) investigating the links between tobacco use and cancer at one of six sites: the lung, oral cavity, larynx, oesophagus, urinary bladder and stomach. Each epidemiological study reviewed is classified by the decision-maker according to whether or not it suggests that tobacco use causes cancer at the site of interest. The decision-maker's beliefs concern the population proportion of such studies that would conclude that tobacco use causes cancer at the site, were it possible to observe all of the studies in the population. Prior to observing any studies, the decision-maker believes that tobacco use is not harmful to the site and therefore expects that few studies in the population will conclude that it is. As time passes, the decision-maker reviews newly-published articles and classifies each according to whether it suggests that tobacco use is harmful, beneficial, or neutral, or whether the study results are inconclusive or unclear. Beliefs are updated using this new information and, from time to time, are combined with evidence from other sources (such as the results of experimental work, chemical and pharmacological analyses of tobacco and its effects), in an assessment of whether or not the evidence is strong enough to declare publicly that the relationship between tobacco use and cancer of the site in question is causal.

3.1 The model

Consider the population of all epidemiological studies investigating the relationship between tobacco use and cancer in one of the six sites. Let $p \in [0,1]$ be the population proportion of studies that conclude that tobacco use is harmful. Let $P_t \in [0,1]$ be a Beta random variable ($P_t \sim \text{Be}(a_t, b_t)$, $a_t, b_t \in \mathbb{R}_+$) defined at each point $t \in \mathbb{T}$ of the time horizon, where $\mathbb{T}$ is the set $\{t_0, t_1, \ldots, T\}$ and $T < \infty$. t_0 is a point in time prior to the date of

[8] The authors consider only articles written in English and exclude 'all Scandanavian, British and Canadian articles'.

publication of the first epidemiological study reviewed by the decision-maker and T is the final point in the time horizon (in our model the ts represent years and $t_0 = 1919$ and $T = 1963$; 1920 was the year in which the first case-control study for our series (looking at oral cancer (Broders, 1920)) was published).

Prior beliefs are defined by choosing suitable values for a_{t_0} and b_{t_0}, yielding the following density f for P_{t_0}:

$$f(P_{t_0}) \propto p^{a_{t_0}-1}(1-p)^{b_{t_0}-1}. \tag{1}$$

For each $t \in \mathbb{T}$, the decision-maker observes a total of $n_t \in \mathbb{Z}_+$ new studies meeting its inclusion criteria, of which $x_t \in \{0, \ldots, n_t\}$ are classified by the decision-maker as suggesting that tobacco use is harmful.

The decision-maker, in classifying a study, can take into account the results of 'classical' statistical hypothesis tests carried out as part of the study, together with other factors such as an assessment of the overall quality of the study's design, the competence of the study's investigators, the conclusions drawn by the investigators, and so on. Assuming a prior position that tobacco is not harmful, the decision-maker knows that, even if it is truly the case that tobacco is not harmful, some studies might incorrectly conclude that it is, perhaps because a hypothesis test in the study has made a Type I error. The decision-maker is therefore aware that, in the population of studies, it would be possible to observe a low proportion of studies concluding that tobacco use is harmful, even if it is not. However, the decision-maker is also aware that it is possible that tobacco use *is* harmful, in which case the decision-maker would expect to observe a higher proportion of studies classified as 'tobacco harmful' in the population. To reflect these beliefs, the prior for p will be centered around a low, but non-zero, value (implying that $E[P_{t_0}]$ is low) but will assign a non-zero probability to p taking higher values. The strength of the decision-maker's prior beliefs will be reflected in $\text{var}(P_{t_0})$. Appendix A provides a more detailed interpretation of the 'meaning of 'p" under the two scenarios 'tobacco not harmful/neutral' and 'tobacco harmful.'

In year t_1, the decision-maker observes a sample of n_{t_1} studies from the population, of which x_{t_1} are classified as suggesting that tobacco use is harmful.[9] Assuming that the sample consists of random and independent draws from the population, the probability of classifying x_{t_1} studies as 'tobacco harmful', from a total of n_{t_1} new studies, is of the form:

$$g(x_{t_1}|P) \propto p^{x_{t_1}}(1-p)^{n_{t_1}-x_{t_1}}. \tag{2}$$

Bayes' theorem for the posterior distribution of beliefs is:

$$f(P_{t_1}|x_{t_1}) \propto g(x_{t_1}|p)f(P_{t_0}), \tag{3}$$

[9]We note that, since $n_t \in \mathbb{Z}_+$, the number of new studies published in any year can be zero.

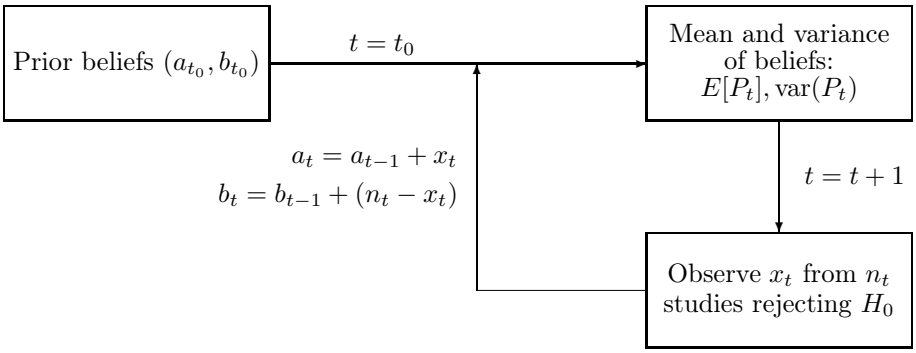

FIGURE 2: Updating process for beliefs.

implying, using (1) and (2), that posterior beliefs are such that:

$$f(P_{t_1}|x_{t_1}) \propto p^{a_{t_0}+x_{t_1}-1}(1-p)^{b_{t_0}+(n_{t_1}-x_{t_1})-1}.$$

Hence, at the posterior stage, $P_{t_1} \sim \text{Be}(a_{t_0} + x_{t_1}, b_{t_0} + (n_{t_1} - x_{t_1}))$.

The process of updating beliefs can be represented using a simple set of difference equations and is shown in Figure 2. Given starting values for a_{t_0} and b_{t_0}, a_t and b_t will evolve according to:

$$a_t = a_{t-1} + x_t, \tag{4}$$
$$b_t = b_{t-1} + (n_t - x_t), \tag{5}$$

$t = t_0 + 1, \ldots, T - 1$. The standardised likelihood for beliefs at each $t \in \mathbb{T}$ is then:

$$f(P_t|x_t) = \frac{\Gamma(a_t + b_t)}{\Gamma(a_t)\Gamma(b_t)} p^{a_t - 1}(1-p)^{b_t - 1},$$

where Γ is the gamma function. Expressions for the mean and variance of P_t are given by:

$$E[P_t] = \frac{a_t}{a_t + b_t}, \tag{6}$$
$$\text{var}(P_t) = \frac{a_t b_t}{(a_t + b_t)^2(a_t + b_t + 1)}. \tag{7}$$

The idea is that, when a causal relationship does exist between tobacco use and cancer of a particular site, there arrives a point at which the 'weight of evidence' (reflected in $E[P_t]$), combined with a particular degree of certainty (reflected in $\text{var}(P_t)$) and other corroborative evidence, 'persuades'

the decision-maker to declare the relationship between tobacco use and the cancer 'causal'.

The major differences between Chern et al.'s model and ours is that, firstly, Brown and Schrader's 'cholesterol index' does not appear to have limited its inclusion criteria to epidemiological studies.[10] Secondly, Chern et al. interpret their model's outputs to be beliefs concerning diet and disease relationships held by the population whereas, in our model, a decision-maker's beliefs are limited to the population proportion of epidemiological studies concluding that tobacco use is harmful.

For the rest of this section we describe the model for one cancer site. The methods described are applied to all sites in models for: (a) the T.I.R.C. and (b) the U.S. Public Health Service and related groups.

3.2 Prior beliefs

From reading the history of the tobacco story, it is clear that there was initial resistance in the medical, scientific and wider community, to the idea that tobacco use could harm human health. We account for this in setting prior beliefs to reflect the decision-maker's beliefs that few studies in the population conclude that tobacco use is harmful. To reflect the fact that the epidemiological studies reviewed all apply 'classical' statistical hypothesis tests, we choose $E[P_{t_0}] = 0.0375$, reflecting the proportion of studies one would expect incorrectly to conclude that tobacco is harmful if the average probability of making a Type I error within the population of all studies equals 3.75%. $E[P_{t_0}] = 0.0375$ implies, from (6), that $a_{t_0} = 3b_{t_0}/77$. We choose $b_{t_0} = 38$, implying that $a_{t_0} = 114/77$ and the distribution for P_{t_0} is such that there is a probability of 0.95 that p lies within the interval [0.003,0.11]. Appendix A discusses this choice of prior beliefs in more detail.

3.3 Updating of beliefs

The sources of data used to update beliefs are taken from tables 2 to 9 of Forster et al. (2006). These list the major prospective and retrospective studies investigating the relationship between tobacco use and the six cancer sites considered in this paper and identified by Forster et al. as having been reviewed by the T.I.R.C. and/or the U.S. Public Health Service and related groups during the 1950s and early 1960s.[11]

Starting with the prior values for a_{t_0} and b_{t_0}, for each organisation and for all $t \in \mathbb{T}$, x_t and n_t were calculated from the tables and substituted into (4) and (5), which were, in turn, used to calculate the mean and variance of

[10] The index appears to comprise all articles dealing with the 'human and clinical implications' of cholesterol with regard to heart disease or arteriosclerosis.

[11] Full details on how the tables were compiled may be found in Forster et al. (2006).

P_t using (6) and (7).[12] Two sets of results are presented for the six cancer sites. The first set deals exclusively with the studies reviewed in the U.S. Surgeon General's Advisory Committee's report of 1964. For each of the six cancer sites, we plot the evolution of the mean and variance of P according to the year in which the studies reviewed by the Advisory Committee were published. n_t therefore represents the total number of studies published in year t and x_t the number which the Advisory Committee classified as 'tobacco use harmful' in its 1964 report. In this section, the outputs of the model show how the Committee's beliefs would have evolved had it assessed the evidence considered in 'Smoking and Health' as that evidence emerged. For each cancer site, we compare the values $E[P_{1963}]$ and $\text{var}(P_{1963})$ with the Advisory Committee's evaluation of the evidence presented in its report. This set of results aims to show, firstly, in graphical format, how the epidemiological evidence evolved over time for each of the six cancer sites. Secondly, it is used to consider how 'well' the model performs, in terms of the degree to which the assessments of the evidence made by the Advisory Committee in 1964 match the results of the model.

In the second set of results, we consider the perspectives of the U.S. Public Health Service and related groups and the T.I.R.C. between 1950 and 1963. The T.I.R.C. series are plotted according to the dates at which we know the evidence was reviewed by the T.I.R.C. (in either the Catalog (1955) or the Current Digest (yearly from 1956 onwards)). We also consider how the T.I.R.C. interpreted the evidence for lung cancer in the annual reports of its Scientific Director. For the U.S. Public Health Service and related groups, the series are plotted according to the dates at which the service reviewed the evidence and released public statements evaluating it. This corresponds to the years 1957 (report of the Study Group on Smoking and Health and statement of the Surgeon General), 1959 (second statement of the Surgeon General) and 1964 (year of publication of 'Smoking and Health'). The focus in the second set of results is therefore on the timing of reviews of the evidence by the T.I.R.C. and the Public Health Service and related groups and the degree to which outputs of the models for the six cancer sites match the public statements of the organisations over time.

Before turning to the results, we note that we are not directly modelling the Bayesian approach to assessing the probability that the null hypothesis

[12] For the T.I.R.C. series for the years prior to 1956, x_t is given by the total number of 'C' (tobacco harmful) classifications in Forster et al.'s tables. For the T.I.R.C. series for 1956 and beyond, x_t is the total number of 'Y' classifications recorded in the tables. Studies noted in Forster et al.'s tables as not being clear retrospective studies in the Current Digest summary were excluded from the series. The classifications of the studies in the possession of the U.S. Public Health Service and related groups were taken from its public statements of 1957 and 1959 (lung cancer only) and Forster et al.'s tables relating to the Advisory Committee's report of 1964. In rare cases where a study was interpreted as being 'unclear' in its conclusions (a 'B' (Catalog) or 'U' (all other sources) in Forster et al.'s tables), we classified it as 'tobacco not harmful'.

is true given a set of data (i.e. $P(H_0 \text{ true} \mid \text{data})$). Rather, the estimation concerns the population parameter p which, in turn, forms part of a wider body of evidence used by the decision-maker to assess the relationship between tobacco use and cancer. There will be no 'threshold' level of p beyond which a relationship is automatically declared as being 'causal'. Further, as discussed in the appendix, when tobacco is harmful, the population value of p depends on, amongst other factors, the distribution of the power of the hypothesis tests within the population of studies to detect a true difference. This is, in turn, determined by factors such as the sample size of the studies and the degree to which the disease is multi-causal in nature (other things being equal, the more the presence of a disease can be 'explained' by tobacco use, the greater will be the power of hypothesis tests to detect a true difference).

4 Results

4.1 Results for the Surgeon General's Advisory Committee's report of 1964

In this section of the paper, our models use as inputs the year of publication of the articles reviewed by the Surgeon General's Advisory Committee that are listed in Tables 2 to 9 of Forster et al. (2006). Table 1 compares, for the Surgeon General's Advisory Committee, the mean and variance of the distribution of beliefs regarding p in 1963 to those of prior beliefs (in 1919). Plots of how the mean and variance of p evolve over time for each cancer site, together with details of the Committee's evaluation of the evidence within its possession in 1963, are presented in Figures 3 to 5. To aid comparability, we plot each series against a common time line, using 1919 as the starting point.

Table 1 shows that $E[P_{1963}]$ is higher than $E[P_{1919}]$ for all cancer sites except stomach cancer, reflecting the increase in the 'weight' of epidemiological evidence linking tobacco use to cancer for these sites during the period in question. Viewed in terms of the expected values, the table shows the weight of evidence to be strongest for lung cancer, followed by cancer of the oral cavity,[13] laryngeal cancer, oesophageal cancer and cancer of the urinary bladder. The variance is higher in 1963 than in 1919 for all five of these

[13]For oral cancer, the Surgeon General's Advisory Committee broke down the results of studies to consider cancer at specific sites of the oral cavity and its relation to specific forms of tobacco use (pipe smoking, chewing and so on). For our model, we considered a study's results to support a causal link between tobacco use and oral cancer as long as the Advisory Committee classified the study as suggesting that tobacco use caused cancer at least one site within the oral cavity. For some studies, although such a classification was made for one site in the oral cavity, links between tobacco use and cancer at other sites in the oral cavity were either classified by the Advisory Committee as associations 'of doubtful significance' or 'absent or not significant'.

Cancer site	$E[P_t]$	$\text{var}(P_t)$	Overall assessment
Starting values for all sites ($t = 1919$)	0.0375	0.0009	
Values of series ($t = 1963$)			
Lung	0.4966	0.0033	Causal
Oral cavity	0.3272	0.0038	Causal (lip only)
Larynx	0.1916	0.0031	Causal
Oesophagus	0.1429	0.0026	Association
Urinary bladder	0.1260	0.0025	Association
Stomach	0.0340	0.0007	No relationship

TABLE 1: Surgeon General's Advisory Committee: means ($E[P_t]$) and variances ($\text{var}(P_t)$) for the distribution of beliefs for each cancer site in 1919 (prior) and 1963, together with the Committee's judgment as to the nature of the relationship between tobacco use and cancer at each site.

cancer sites, reflecting the fact that prior beliefs about p have been contradicted by the sample evidence. For stomach cancer, $E[P_t]$ and $\text{var}(P_t)$ vary little between 1919 and 1963, reflecting the fact that none of the four studies reviewed by the Advisory Committee were classified as suggesting that a causal link existed between tobacco use and stomach cancer; the sample evidence supports prior beliefs and hence mean and variance change little.

Comparing the way the model ranks the six sites to the overall assessment regarding the relationship between tobacco use and cancer at each site as made by the Advisory Committee, table 1 shows a reasonable match. The Committee declared that a 'causal' relationship existed for three of the sites (lung, oral cavity (lip only) and larynx), that associations existed for two of them (oesophagus and urinary bladder) and that no relationship had been established between tobacco use and stomach cancer. The model's ranking in table 1 fits these assessments.

Consider now how the series evolve over time. Figure 3 shows that, for both lung and oral cancer, $E[P_t]$ is non-decreasing throughout the period in question. This reflects the fact that, for both sites, every study reviewed by the Surgeon General's Advisory Committee was classified as 'tobacco harmful'. For oral cancer, the first change in the mean and variance occurs in 1920, with the publication of Broders (1920). Thereafter, sporadic publication of studies increases the mean during the rest of the 1920s, 1930s and 1940s, before more frequent publication of studies is observed during the 1950s. As the total number of studies classified as 'tobacco harmful' increases, the marginal impact of a study concluding that tobacco is harmful on $E[P_t]$ is strictly positive but strictly decreasing.[14] This can be observed by comparing the increment in the mean between years when one additional

[14]Differentiating (6) with respect to a yields $dE[P]/da = b/(a+b)^2 > 0; d^2E[P]/da^2 = -2b/(a+b)^3 < 0$.

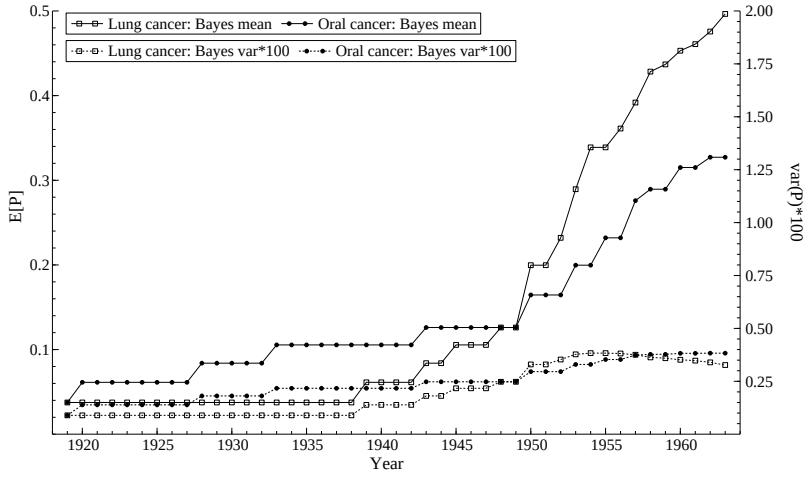

FIGURE 3: Lung and oral cancer series for Surgeon General's Advisory Committee: mean and 100*variance of distribution of beliefs regarding p, 1919 - 1963.

Advisory Committee's assessment of the evidence for each site:
Lung cancer - 'Cigarette smoking is causally related to lung cancer in men; the magnitude of the effect of cigarette smoking far outweighs all other factors. The data for women, though less extensive, point in the same direction.'
Cancer of the oral cavity - 'The causal relationship of the smoking of pipes to the development of cancer of the lip appears to be established. Although there are suggestions of relationships between cancer of other specific sites of the oral cavity and the several forms of tobacco use, their causal implications cannot at present be stated.'

publication is observed, for example between 1919 and 1920 and 1961 and 1962. For oral cancer, the variance of P increases throughout the time period, reflecting the increased uncertainty associated with classifying studies suggesting that tobacco is harmful, having taken a prior position that it is not.

For lung cancer the story is similar, although it is compressed into a shorter time-span and the weight of evidence suggesting that tobacco use is harmful exceeds that for oral cancer from 1950 onwards. The first change in the lung cancer series occurs in 1939, with the publication of Müller (1939). Sporadic studies are observed during the 1940s, with the main body of evidence accumulating during the 1950s. In 1949, beliefs regarding P are identical for lung and oral cancer, reflected in the identical values for $E[P_{1949}]$ and $\text{var}(P_{1949})$. Thereafter, the lung cancer series has a higher mean and lower variance than the oral cancer series. $\text{var}(P_t)$ for lung cancer rises at first but

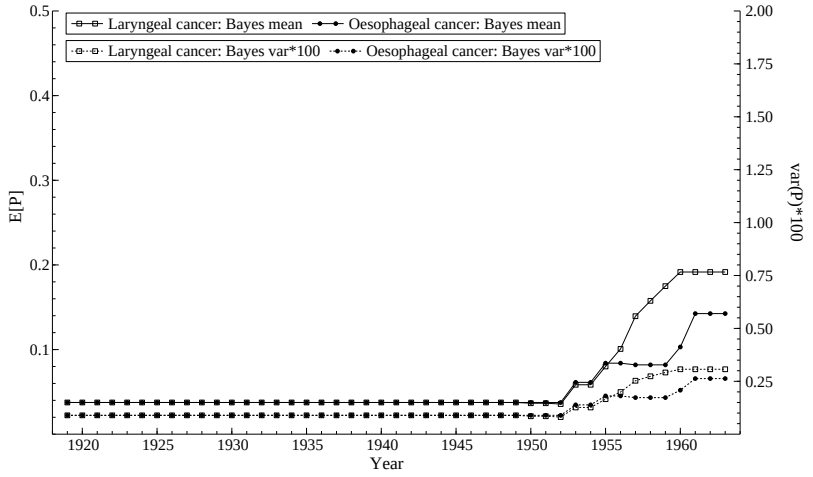

FIGURE 4: Laryngeal and oesophageal cancer series for Surgeon General's Advisory Committee: mean and 100*variance of distribution of beliefs regarding p, 1919 - 1963.

Advisory Committee's assessment of the evidence for each site:
Laryngeal cancer - 'Evaluation of the evidence leads to the judgment that cigarette smoking is a significant factor in the causation of laryngeal cancer in the male.'
Oesophageal cancer - 'The evidence on the tobacco-esophageal relationship supports the belief that an association exists. However, the data are not adequate to decide whether the relationship is causal.'

falls from 1955 onwards, after the sample information becomes strong enough to reduce the uncertainty introduced by classifying studies concluding that tobacco use is harmful.

Figures 4 and 5 show that, for laryngeal, oesophageal and urinary bladder cancer, the story is similar. However, the increments for these series are compressed into an even smaller time frame and the weight evidence by 1963 is less strong. Figure 5 shows how the mean and variance for stomach cancer are virtually invariant over time: four studies reported, in 1946, 1957, 1960 and 1961, and all were classified by the Advisory Committee as showing that tobacco use was not harmful, which is consistent with prior beliefs.

4.2 Results for the T.I.R.C. and the Surgeon General's Advisory Committee compared

Results comparing $E[P_{1963}]$ and $\text{var}(P_{1963})$ based on the evidence collected by the U.S. Surgeon General's Advisory Committee and that collected by the T.I.R.C. are shown in table 2. For lung cancer, the means and variances

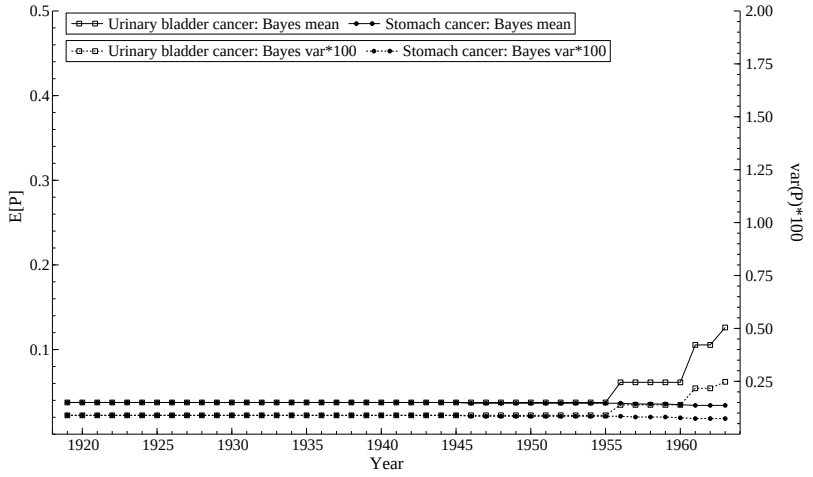

FIGURE 5: Urinary bladder and stomach cancer series for Surgeon General's Advisory Committee: mean and 100*variance of distribution of beliefs regarding p, 1919 - 1963.

Advisory Committee's assessment of the evidence for each site:
Urinary bladder cancer - 'Available data suggest an association between cigarette smoking and urinary bladder in the male but are not sufficient to support a judgment on the causal significance of this association.'
Stomach cancer - 'No relationship has been established between tobacco use and stomach cancer.'

for each organisation are almost identical in 1963. The T.I.R.C.'s slightly lower mean and higher variance are accounted for by the fact that it had reviewed slightly fewer studies in the Catalog and Current Digest than the Advisory Committee and that one of the studies reviewed by the T.I.R.C. was classified as 'B, unclear' in the Catalog (Watson, 1950) and therefore as 'tobacco not harmful' in the Bayesian model. For oral cancer, the lower mean and variance for the T.I.R.C. in 1963 are accounted for mainly by the low number of studies published prior to August 1955 that had been summarised in the Catalog. For the other four cancer sites, the T.I.R.C.'s mean and variance are higher than those of the Advisory Committee.[15]

[15]For laryngeal cancer this is because the Surgeon General's Advisory Committee classified two studies (Schrek et al. (1950) and Valko (1952)) as being unclear in their conclusions. The T.I.R.C. classified Schrek et al. as 'C, tobacco harmful' in the Catalog and did not review Valko. For oesophageal cancer, the T.I.R.C. series is higher because its series comprises more studies, namely Stocks (1957), Pernu (1960), Shanta and Krishnamurthi (1963) and Levin (1963). For urinary bladder and stomach cancer, the T.I.R.C.'s mean and variance are higher because it reviewed and classified as harmful one more study than the Advisory Committee (Levin, 1963 and Pernu, 1960 respectively).

	Surgeon General's Advisory Committee		T.I.R.C.	
Cancer site	$E_t[P]$	$\text{var}_t(P)$	$E[P_t]$	$\text{var}(P_t)$
Starting values for all sites ($t = 1919$)	0.0375	0.0009	0.0375	0.0009
Values of series ($t = 1963$)				
Lung	0.4966	0.0033	0.4619	0.0034
Oral cavity	0.3272	0.0038	0.2619	0.0037
Larynx	0.1916	0.0031	0.2320	0.0035
Oesophagus	0.1429	0.0026	0.1645	0.0030
Urinary bladder	0.1260	0.0025	0.1457	0.0027
Stomach	0.0340	0.0007	0.0570	0.0012

TABLE 2: U.S. Surgeon General's Advisory Committee and T.I.R.C.: comparison of means and variances for the distribution of beliefs regarding p for each cancer site in 1963.

Figures 6 to 11 compare the evolution of the series for the T.I.R.C. and the U.S. Public Health Service and related groups between 1950 and 1963. The T.I.R.C. series are plotted according to the year at which the studies appeared in the Catalog (1955 and for lung, oral and laryngeal cancer only) and the Current Digest (yearly from 1956 onwards and for all cancer sites). The series for the U.S. Public Health Service and related groups are plotted according to the date at which the evidence was reported and assessed in a public statement. For lung cancer, this corresponds to the years 1957 (Study Group on Smoking and Health, 1957 and Burney, 1958), 1959 (statement of Burney (1959)) and 1964 ('Smoking and Health'). The other five sites do not appear to have been comprehensively reviewed by the service until 1964, which accounts for the single increments in the Public Health Service series in Figures 7 to 11.

For the lung cancer series, Figure 6 also compares the assessments of the evidence regarding smoking and lung cancer made by the T.I.R.C. in its annual reports of the Scientific Director and the U.S. Public Health Service and related groups, in 1957, 1959 and 1964. A plot of the Beta distributions for the U.S. Public Health Service and related groups at the prior stage, 1957, 1959 and 1963 is shown in figure 12.

Figure 6 shows how, for lung cancer, the information contained within the T.I.R.C.'s Catalog of 1955 - every study but one reviewed in this document was given a 'C, tobacco harmful' classification - caused $E[P_{1955}]$ to jump (from its prior value of 0.0375 to 0.2842) and the variance to rise (0.0009 to 0.0037). Thereafter, the mean increases steadily, reflecting the fact that the T.I.R.C. was swift to make summaries of studies available in the Current Digest. The U.S. Public Health Service series changes first in 1957, with the publication of the Study Group's report on smoking and health. It can be seen that the study group's report included more studies classified as

'tobacco harmful', since the 1957 mean of this series is higher than that of the T.I.R.C.. The series increments again in 1959, when the Surgeon General added an additional six studies to the evidence base, and, finally, again in 1963, when twelve further studies were added by the Advisory Committee's report.

The quotes in Figure 6 demonstrate the striking difference in the interpretation of the evidence available to the organisations. The Study Group and the Surgeon General were suggesting a causal relationship for excessive smoking and lung cancer in 1957 (phrases used include 'beyond reasonable doubt' and 'the weight of evidence is increasingly pointing in one direction'). In 1959, the Surgeon General stated that the weight of evidence indicated that smoking (rather than excessive smoking) was the 'principal factor' in the increased incidence of lung cancer. By 1964 the relationship was defined as being unequivocally 'causal'. These conclusions are not at odds with the U.S. Public Health Service and related groups' series plotted in Figure 6, although it must be stressed the statements reflected an assessment of all of the evidence available to these organisations, not solely the epidemiological evidence which provides the input to our model. The T.I.R.C.'s statements contrast sharply with the statements of the Public Health Service and related groups as well as the series from the T.I.R.C.'s own model. The Scientific Director's 1964 statement - that evidence suggesting that cigarettes exercise a direct, carcinogenic effect on man had not been forthcoming - is of particular note. The possible reasons for this are addressed in the discussion.[16]

Finally, we note the T.I.R.C. series for the other five cancer sites in Figures 7 to 11. We did not find any detailed review or discussion of the evidence relating to these cancer sites in the annual reports of the T.I.R.C.'s Scientific Director for the period in question.

[16] A more detailed coverage of the annual reports of the T.I.R.C.'s Scientific Director are presented, year by year between 1956 and 1964/5, in Forster et al. (2006).

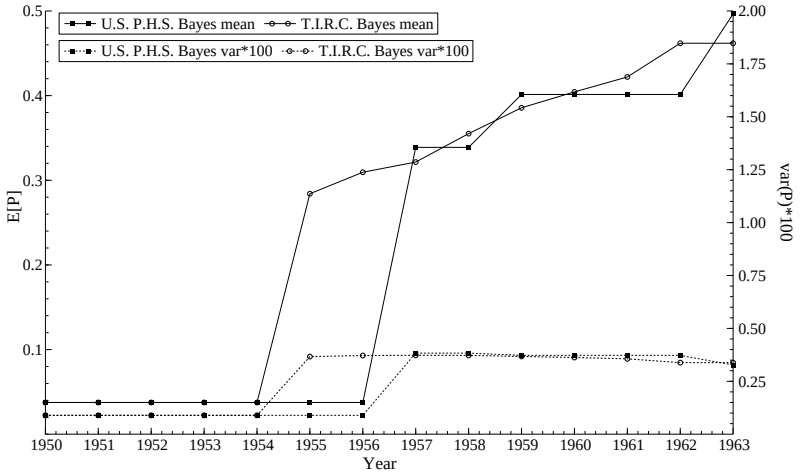

FIGURE 6: Lung cancer series for U.S. Public Health Service and related groups (U.S. P.H.S.) and T.I.R.C. compared: mean and 100*variance of distribution of beliefs regarding p, 1950 - 1963. Assessments of the evidence:

1957
Surgeon General's Study Group: '[t]he sum total of scientific evidence establishes beyond reasonable doubt that cigarette smoking is a causative factor in the rapidly increasing incidence of human epidermoid cancer of the lung';
Surgeon General Burney: 'the weight of evidence is increasingly pointing in one direction: that excessive smoking is one of the causative factors in lung cancer.'
T.I.R.C.'s Annual Report of the Scientific Director: 'such data would provide clear evidence of the fact that the outward and visible habit of heavy smoking is a reflection of such a wide and varied gamut of internal disturbances and unbalances that its possible specific, *causative* value, becomes reduced almost to an absurdity.'

1959
Surgeon General Burney: 'the weight of evidence at present implicates smoking as the principal factor in the increased incidence of lung cancer ... cigarette smoking particularly is associated with an increased chance of developing lung cancer.'
T.I.R.C.'s Annual Report of the Scientific Director: 'growing support for the point of view that the statistical association claimed by various studies has an explanation or explanations that may still not be apparent from our present knowledge.'

1964
Surgeon General's Advisory Committee: 'Cigarette smoking is causally related to lung cancer in men; the magnitude of the effect of cigarette smoking far outweighs all other factors. The data for women, though less extensive, point in the same direction.'
T.I.R.C.'s Annual Report of the Scientific Director: 'over the years, however, evidence to support the thesis that cigarettes exercise a direct carcinogenic effect on man has not been forthcoming.'

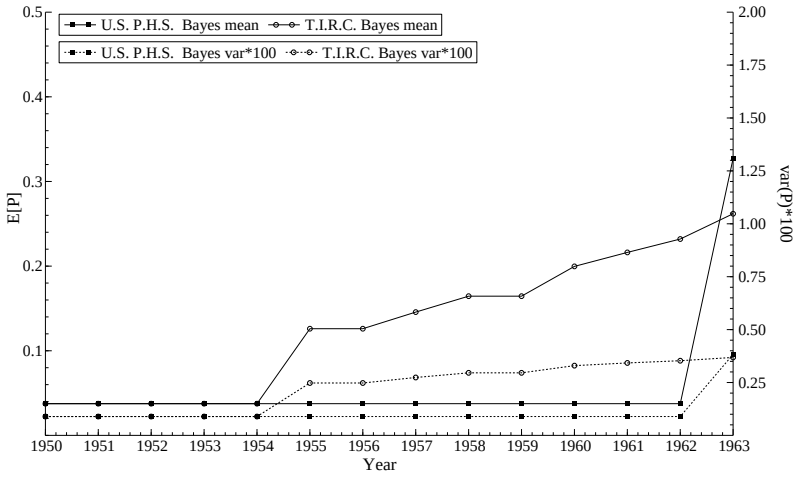

FIGURE 7: Oral cancer series for Surgeon General's Advisory Committee and T.I.R.C. compared: mean and 100*variance of distribution of beliefs regarding p, 1950 - 1963.

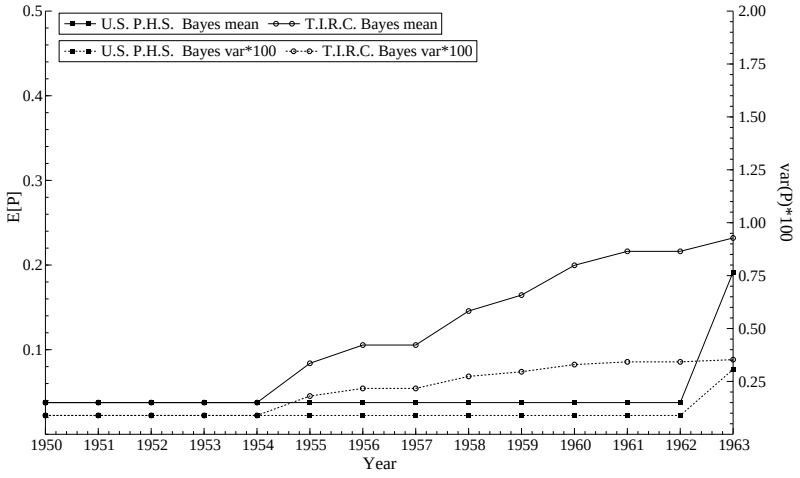

FIGURE 8: Laryngeal cancer series for Surgeon General's Advisory Committee and T.I.R.C. compared: mean and 100*variance of distribution of beliefs regarding p, 1950 - 1963.

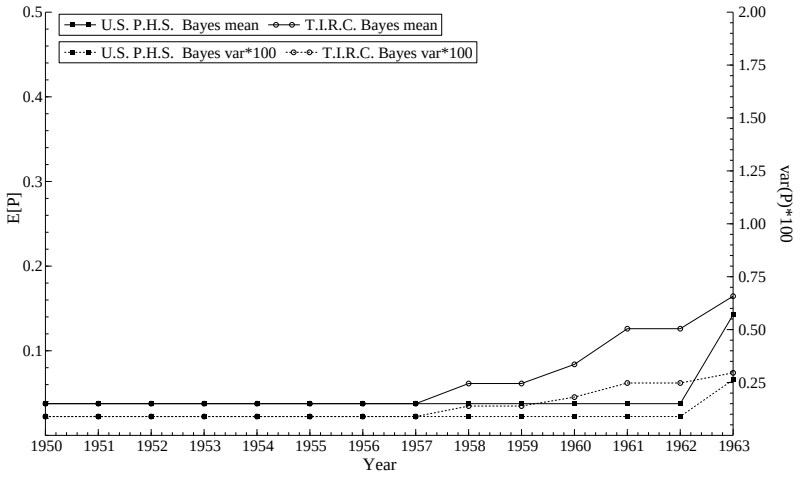

FIGURE 9: Oesophageal cancer series for Surgeon General's Advisory Committee and T.I.R.C. compared: mean and 100*variance of distribution of beliefs regarding p, 1950 - 1963.

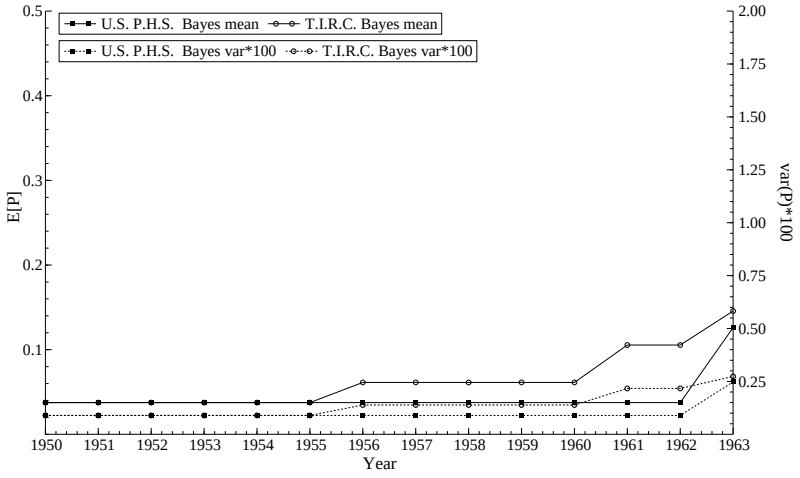

FIGURE 10: Urinary bladder cancer series for Surgeon General's Advisory Committee and T.I.R.C. compared: mean and 100*variance of distribution of beliefs regarding p, 1950 - 1963.

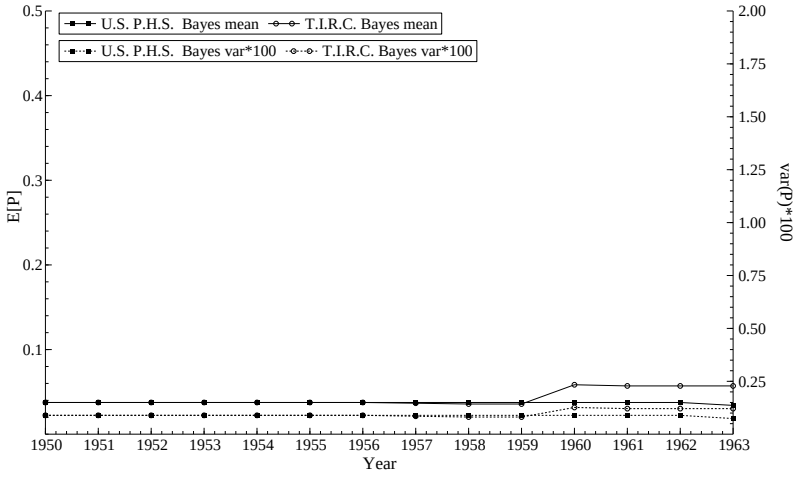

FIGURE 11: Stomach cancer series for Surgeon General's Advisory Committee and T.I.R.C. compared: mean and 100*variance of distribution of beliefs regarding p, 1950 - 1963.

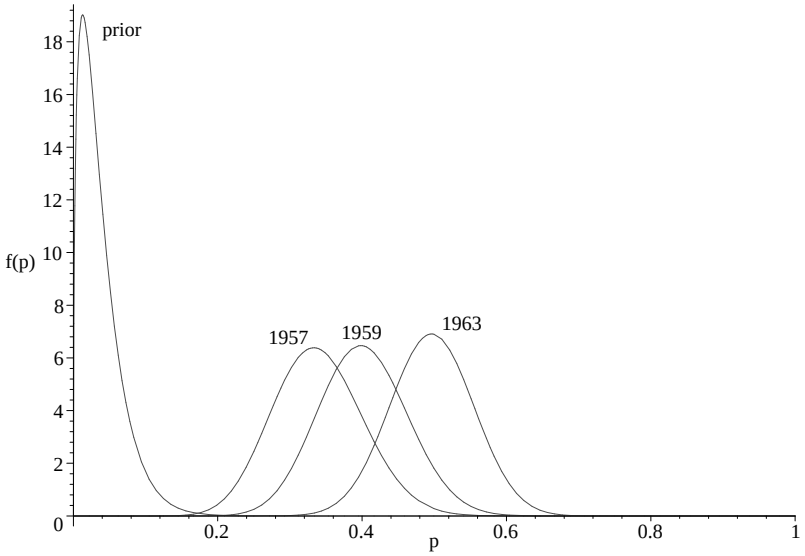

FIGURE 12: Evolution of Beta probability density function for beliefs regarding p for lung cancer, U.S. Public Health Service and related groups, prior, 1957, 1959 and 1963.

5 Discussion

Viewed in terms of the ranking presented in table 1, the results of the Bayesian model for the Surgeon General's Advisory Committee are not at odds with the overall assessments of the evidence made by it. Furthermore, for lung cancer, the results plotted in figure 6 for the Study Group, Surgeon General and Advisory Committee are reasonably consistent with the way in which they assessed the accumulating body of evidence between 1957 and 1963.

What are radically different are the assessments of the evidence in the annual reports of the T.I.R.C.'s Scientific Director, when viewed in the light of the output from the T.I.R.C. models. Firstly, for all sites other than lung cancer, the reports failed to make any significant reference to the evidence base that was within the T.I.R.C.'s possession. Secondly, for lung cancer, the differences in the assessment of the evidence are especially striking when one considers that, by 1963, the T.I.R.C.'s series for lung cancer is virtually identical to that of the U.S. Public Health Service and related groups. Although the annual reports of the T.I.R.C.'s Scientific Director made repeated references to the statistical associations between smoking and lung cancer that had been reported within the epidemiological literature, they were not, on their own or in combination with other evidence collected by the T.I.R.C., sufficient to persuade the Scientific Director to declare the relationship between tobacco use and lung cancer 'causal'.

What factors could account for the failure of the model for lung cancer to match the statements of the T.I.R.C.? The first possibility is that, since the epidemiological evidence was only one source of evidence used to assess the causal relationship between tobacco use and lung cancer, the T.I.R.C. could simply have attached far less weight to the epidemiological evidence when assessing the overall body of evidence. This possibility is supported in the annual reports of the Scientific Director, where frequent references are made to the limitations of epidemiological evidence and statistical associations in making judgments about causality. This is also the line of argument that has been made by Talley et al. (2004) and Parascandola (2004) who use, amongst others, the views of Fisher and Berkson as well as statements of Dr. Lewis Robbins, Chief of Cancer Control Program, U.S. Public Health Service, to support the point.[17]

The line of argument becomes confused, however, when one considers that epidemiological evidence was, at times, used by the T.I.R.C. to make judgments about causality. Forster et al. (2006) cite the example of the Framingham study, a prospective epidemiological study, to which Little referred in his 1964/5 annual report: 'cigarette smokers have no greater inci-

[17]For example, Talley et al. quote Robbins in 1962: 'from the first hint of a relationship between smoking and cancer there was controversy, and that controversy continues' (Robbins, 1962).

dence of angina pectoris than non-smokers. Since this syndrome is regarded as most often being a chronic manifestation of coronary artery sclerosis, this finding suggests that smoking does not cause or accelerate such sclerosis ... at the present time therefore, the weight of evidence is against the concept that either nicotine or smoking influences the atherosclerotic process.' Forster et al. note that, prior to 1964, the T.I.R.C.'s Current Digest had reported other evidence from the Framingham study suggesting that heavy cigarette smokers experienced a threefold increase in the incidence of myocardial infarction, but that this was not recorded by Little in any annual report.

Secondly, even if the T.I.R.C. did give some weight to the epidemiological evidence, it is clear that the prior beliefs of the T.I.R.C. and those of the U.S. Public Health Service and related groups need not have been identical: the T.I.R.C.'s prior beliefs could have differed in terms of both their location and spread. We believe that it is unrealistic to argue that, at the prior stage, the T.I.R.C. could have expected fewer than around 3.75% of studies incorrectly to reject a true null, given that the majority of hypothesis tests in the studies appear to have applied hypothesis tests using 5% significance levels (one or two-tailed). However, it is quite possible that the variance of the T.I.R.C.'s distribution of beliefs was smaller than that of the U.S. Public Health Service and related groups.

Using techniques similar to those of Degroot and Schervisch (2002), we can ask what strength of prior beliefs around $E[P_{1919}]$ could have led to the T.I.R.C. drawing a conclusion that an association, but not a causal relationship, existed between tobacco use and lung cancer, if its assessment criteria were similar to those of the Surgeon General's Advisory Committee. The Advisory Committee found the evidence within its possession to have been strong enough to declare a causal relationship to exist between tobacco use and laryngeal cancer ($E[P_{1963}] = 0.1916, \text{var}(P_{1963}) = 0.0031$) but not oesophageal cancer ($E[P_{1963}] = 0.1429, \text{var}(P_{1963}) = 0.0026$). Therefore, a simple approach is to ask how 'tight' prior beliefs for the T.I.R.C. would have had to have been such that, for its beliefs in 1963, 0.1429 ('association' declared) $< E[P_{1963}] <$ 0.1916 ('causality' declared). Given that, for lung cancer, the T.I.R.C. classified 32 studies as 'tobacco harmful' and one as 'tobacco neutral', prior values $a_{t_0} = 9.71$ and $b_{t_0} = 249.16$ would yield $E[P_{t_0}] = 0.0375$ and $E[P_{1963}] = 0.1429$. Prior values $a_{t_0} = 6.26$ and $b_{t_0} = 160.38$ would yield the same expected value for the prior and $E[P_{1963}] = 0.1916$. One way to evaluate the strength of these alternative prior beliefs against tobacco being harmful is to consider the probability that they allow for p to lie above a particular value. For example, for the prior beliefs used in our model, there exists a probability of 0.025 that p lies above 0.114. For the more extreme prior just described (that leading to $E[P_{1963}] = 0.1429$) this probability falls to 0.0000029 and for the less extreme prior it falls to 0.00012.

Finally, this being the argument used by the likes of Glantz et al. (1996) and in a number of legal cases, it could have been that, although privately, there was a growing consensus within the tobacco industry that tobacco use could cause cancer, the T.I.R.C. was reluctant to acknowledge this in public. This line of argument ties in with the 'false controversy' story described in the recent R.I.C.O. judgment and in the petitions lodged by the U.S. states which led to the Master Settlement Agreement of 1998. For example, the petition for the state of Iowa (State of Iowa v. R.J. Reynolds Tobacco Co., et al., 1996) cites an industry document sent in 1958 to the Vice President of Research at Phillip Morris from a company researcher stating that 'the evidence ... is building up that heavy cigarette smoking contributes to lung cancer either alone or in association with physical and physiological factors ...' [par. 77c]. This is not too different from the statements of 1957 made by the Study Group on Smoking and Health and the Surgeon General that are quoted in figure 6 of this paper. And in 1963 the petition cites a memorandum from consultants to Liggett stating: '[b]asically, we accept the inference of a causal relationship between the chemical properties of ingested tobacco smoke and the development of carcinoma, which is suggested in the statistical association shown in the studies of Doll and Hill, Horn and Dorn with some reservations and qualifications ...' [par. 77h]. This is not too different from the assessment made by the U.S. Surgeon General's Advisory Committee that is also quoted in figure 6.

6 Conclusion

This paper has presented a Bayesian model of 'health risk belief' to track the evolution of the 'weight' of epidemiological evidence concerning tobacco harm that was in the possession of the T.I.R.C. and the U.S. Public Health Service and related groups during the mid-twentieth century. Viewed according to the date of publication of the articles reviewed by the Surgeon General's Advisory Committee, we have shown how the epidemiological evidence, measured in a simple, qualitative manner, evolved between 1919 and 1963. We have also seen that the results of the models for the U.S. Public Health Service and related groups are not in disagreement with their public assessments of the evidence but that the T.I.R.C.'s model for lung cancer is not consistent with the assessment of the evidence on lung cancer as expressed in the T.I.R.C.'s annual reports of its Scientific Director between 1956 and 1964/5. Although evidence relating to the other five cancer sites had been reviewed by the T.I.R.C., we did not find any detailed review or discussion of it in the annual reports of the Scientific Director.

In considering the possible reasons for the failure of the T.I.R.C.'s model for lung cancer to fit well the statements expressed in the annual reports of its Scientific Director, we note that C.C. Little was not completely averse

to citing results from epidemiological studies since, on at least one occasion, epidemiological evidence was used to support the case that tobacco was not harmful. Whatever the wider argument about the 'controversy' regarding tobacco harm during the 1950s and 1960s, we find the credibility of the T.I.R.C.'s 'line' on association and causation diminished significantly by such selective use of evidence.

In our discussion, we also considered the importance of the strength of prior beliefs in affecting the outputs of the model, as well as the possible conflict that existed in the tobacco industry between judgments being made in private and statements being made in public. The credibility of the 'strength of prior beliefs' argument rests on the credibility of the T.I.R.C. possessing 'tight' prior beliefs which attach very low probabilities to observing a high population proportion of studies concluding that tobacco is harmful. The credibility of the 'public versus private' argument relies on the credibility and persuasiveness of the views expressed 'privately' by tobacco industry employees and consultants, considered alongside the evidence that was in the possession of the T.I.R.C. and the views that the T.I.R.C.'s Scientific Director was expressing in public.

A Interpretation of p under the null and alternative hypotheses

This section expands on the meaning of the population proportion of studies, p, which conclude that tobacco use is harmful and about which beliefs are modelled using P_t. We assume that the studies in the population test at least one null hypothesis against a one- or two-tailed alternative hypothesis. The focus of attention is on whether or not, considering the hypothesis tests and other evidence from the study, the study is classified by the decision-maker as showing:

- tobacco use is not harmful (which would occur if the study is classified as showing tobacco to be beneficial or neutral to health);
- tobacco use is harmful.

Let there be $i \in 0, \ldots, M$, $M \leq \infty$ studies in the population. The significance level of the hypothesis test for study i is set to be $100\alpha_i\%$, $\alpha \in (0, 1)$, and can vary across studies. We also note that the decision-maker's classification of studies is based not solely on the result of hypothesis tests but also on factors such as an assessment of the overall quality of the study's design, the competence of the study's investigators and so on.

Consider the situation in which tobacco is not harmful. The probability that a hypothesis test carried out by study i incorrectly concludes that tobacco use is harmful is equal to α_i (one-tailed test) or $\alpha_i/2$ (two-tailed test). Assuming that, when reviewing studies, the decision-maker might not base its conclusion solely on the result of a single hypothesis test, and/or that authors might use other information in addition to that provided by hypothesis tests to draw final conclusions about tobacco harm, define a separate random variable, $\Lambda \in (0, 1)$, to represent the distribution of Pr(study incorrectly concludes that tobacco is harmful) within the population, where λ is an individual study's realisation of Λ. p is therefore:

$$p = E[\Lambda] = \int_0^1 \lambda\theta(\lambda)d\lambda,$$

where θ is the probability density function for Λ.

If the null hypothesis is not true, then the probability that study i's hypothesis test correctly rejects the null is equal to the power of the hypothesis test, β_i. Again, the conclusions regarding tobacco harm reached by studies within the population may not be based solely on the results of a single hypothesis test, or might include other information available to the authors, and hence we define the random variable $\Omega \in (0, 1)$ to represent the distribution of Pr(study correctly conclude that tobacco use is harmful) in the population. Let ω be an individual study's realisation of Ω. This probability

will be dependent upon the degree to which the disease is multi-causal in nature (which will affect the positioning of the sampling distribution of the point estimate of effect, such as the odds ratio or the relative risk, under the alternative hypothesis) as well as the study's sample size (which will affect the spread of this sampling distribution). In this scenario, p is the expected value of Ω:

$$p = E[\Omega] = \int_0^1 \omega\epsilon(\omega)d\omega,$$

where ϵ is the probability density function for Ω.

Acknowledgements. Supported by funding from the University of York's Innovation and Research Priming Fund and the Research Application Support Grant of the Department of Economics and Related Studies. We thank Chris Bojke, Stefano Conti, David Epstein and Joe Swierzbinski for their comments. Earlier versions of this paper were presented as a poster to the Royal Statistical Society's annual conference in September 2006 and to the 2006 seminar series of the Business School and Institute of Applied Health Sciences at the University of Aberdeen and we thank participants for their comments. The authors take full responsibility for any errors. Results from Ox v.4.04 (Doornik, 2002) and Maple v. 5.00 (Maplesoft).

References

Almquist et al. v. American Brands, Inc., et al. (1986). District Court of Jefferson County. Deposition of William D. Jenkins.

Austin, K. (1968). Activities of the CTR-USA Library. `http://legacy.library.ucsf.edu/tid/kxn10a00`.

Bowden, S., Forster, M., and Walsh, M. (2006). Corporate social responsibility, science and the tobacco industry in the USA, 1953-1964. University of York.

Broders, A. C. (1920). Squamous-cell epithelioma of the lip. *JAMA*, 74(10):656–664.

Brown, J. B. and Schrader, L. F. (1990). Cholesterol information and shell egg consumption. *American Journal of Agricultural Economics*, 72(3):548–555.

Burney, L. (1958). Statement of July 12, 1957. *Ca, Bulletin of Cancer*, 8(2):44.

Burney, L. (1959). Smoking and lung cancer. A statement of the Public Health Service. *JAMA*, 171:1829–1837.

Chern, W., Loehman, E., and Yen, S. (1995). Information, health risk beliefs, and the demand for fats and oils. *The Review of Economics and Statistics*, 77(3):555–564.

C.T.R. (not dated). A Brief History of the Council for Tobacco Research - U.S.A., Inc. (CTR) Originally Tobacco Industry Research Committee (TIRC). `http://legacy.library.ucsf.edu/tid/xuj3aa00`.

DeGroot, M. H. and Schervish, M. J. (2002). *Probability and Statistics*. Addison Wesley, USA, third edition.

Doll, R. (1998). Uncovering the effects of smoking: historical perspective. *Statistical Methods in Medical Research*, 7:87–117.

Doornik, J. (2002). *Ox. An object-oriented matrix programming language*. Timberlake Consultants Press and Oxford. `www.doornik.com`, London, fourth edition.

European Environment Agency (2001). Late lessons from early warnings: the precautionary principle 1896 - 2000. Copenhagen.

Forster, M., Walsh, M., Bowden, S., Rodgers, M., and Duffy, S. (2006). The state of knowledge regarding tobacco harm, 1920-1964: industry and public health service perspectives. Centre for Historical Economics and Related Research at York, Discussion Paper no. 1, University of York.

Glantz, S. A., Slade, J., Bero, L. A., Hanauer, P., and Barnes, D. E. (1996). *The Cigarette Papers*. University of California Press, Berkeley, `http://ark.cdlib.org/ark:/13030/ft8489p25j/`, first edition.

Haenszel, W., Shimkin, M., and Miller, H. (1956). Tobacco smoking patterns in the United States. Washington: U.S. Department of Health, Education and Welfare, Public Health Service Publication number (PHS) 463.

Hilton, M. (2000). *Smoking in British Popular Culture 1800 - 2000*. Manchester University Press, Manchester, first edition.

Jodie Bullock v. Philip Morris U.S.A., Inc. (2006). Court of Appeal of the State of California, Los Angeles County Super Ct. No. BC249171.

Kluger, R. (1997). *Ashes to Ashes: America's Hundred-Year Cigarette War, the Public Health, and the Unabashed Triumph of Philip Morris*. Random House, New York, first edition.

Levin, M. L. (1963). Smoking and cancer. Retrospective studies and epidemiological evaluation. *Journal of Chronic Diseases*, 16:375–381.

Manning vs. Benson and Hedges Ltd. (2004). High Court of Ireland IECH 316 (30 July 2004), `www.bailii.org/ie/cases/IEHC/2004/316.html`.

Maplesoft. *Maple*. `www.maplesoft.com`.

McTear v. Imperial Tobacco Ltd. (2005). Scottish Court of Session CSOH_69 (31 May 2005), `www.bailii.org`.

Müller, F. H. (1939). Tabakmissbrauch und lungencarcinom. *Z. Krebsforsch*, 49(1):57–84.

National Association of Attorneys General (1998). Master Settlement Agreement. Available on-line at `www.naag.org/backpages/naag/tobacco/msa`.

Parascandola, M. (2004). Two approaches to etiology: the debate about smoking and lung cancer in the 1950s. *Endeavour*, 28(2):81–86.

Parascandola, M., Weed, D. L., and Dasgupta, A. (2006). Two Surgeon General's reports on smoking and cancer: a historical investigation of the practice of causal inference. *Emerging Themes in Epidemiology*, 3(1).

Pernu, J. (1960). An epidemiological study on cancer of the digestive organs and respiratory system. A study based on 7,078 cases. *Ann. Med. Intern. Fenn.*, 49(Suppl. 33):1–117.

Robbins, L. C. (1962). Medical practice and lung cancer. *Minnesota Medicine*, 45:131–36.

Schrek, R., Baker, L. A., Ballard, G. P., and Dolgoff, S. (1950). Tobacco smoking as an etiologic factor in disease. I. Cancer. *Cancer Research*, 10(1):49–58.

Shanta, V. and Krishnamurthi, S. (1963). Further study in etiology of carcinomas of the upper alimentary tract. *British Journal of Cancer*, 17:8–23.

State of Iowa v. R.J. Reynolds Tobacco Co., et al. (1996). Petition.

Stocks, P. (1957). Cancer, incidence in North Wales and Liverpool region in relation to habits and environment. IX. Smoke and smoking.

Stolley, P. D. (1991). When genius errs: R. A. Fisher and the lung cancer controversy. *American Journal of Epidemiology*, 133:416–425.

Study Group on Smoking and Health (1957). Smoking and Health. Joint Report of the Study Group on Smoking and Health. *Science*, 125(3258):1129–1133.

Talley, C., Kushner, H. I., and Sterk, C. E. (2004). Lung cancer, chronic disease epidemiology, and medicine, 1948-1964. *Journal of the History of Medicine and Allied Sciences*, 59(3):329–374.

T.I.R.C. (not dated (b)). Some hints on use of catalog. `http://legacy.library.ucsf.edu/tid/pxn10a00`.

United States of America v. Philip Morris U.S.A., Inc., et al. (2006). United States District Court for the District of Colombia, Civil Action No. 99-2496 (GK) `www.library.ucsf.edu/tobacco/litigation/usvpm/uspm.pdf`.

U.S. Department of Health, Education and Welfare, Public Health Service (1964). *Smoking and Health. Report of the Advisory Committee to the Surgeon General of the Public Health Service*. D. van Nostrand and Co. Inc., Princetown, New Jersey, first edition.

Valko, P. (1952). Smoking and the occurrence of malignant tumors of the larynx. *Cesk Otolaryng*, 1(3):102–105.

Watson, W. L. (1950). Cancer of the lung: consideration of incidence and etiology. *New York Medicine*, 6(12):15–18.